Eurographics Workshop on Visual Computing for Biology and Medicine (2022)
J. Byška and T. Torsney-Weir (Poster Chairs)



# SpaceWalker: a Visual Analytics Approach to Spatial Transcriptomics Data

C. Li[1], T. Höllt[2], B. Lelieveldt[1]

[1]Leiden University Medical Center, the Netherlands
[2]TU Delft, the Netherlands

**Abstract**

*Spatial transcriptomics (ST) enables profiling the expression of hundreds of genes in tissue sections, down to the level of single cells in their tissue environment. For single cells, these high-dimensional (HD) gene expression profiles enable detailed characterization of cell types, cell states, and cell maturation. The spatial cell context enables the study of cell-cell interactions, tissue architecture, and cell development and migration in the tissue. Various computational approaches have been developed to extract information from either spatial domain or gene expression domain individually. However, integrative biological interpretation of HD single cell and spatial data spaces remains challenging. The relationship between HD single-cell data, spatial location and similarity embedding has not been fully explored. In this work, we present SpaceWalker, an interactive visual analytics tool for exploring the spatial structure of ST data, while linking it to (developmental) cell phenotype information computed from the HD gene expression profiles. Specifically, we explored approaches where the user is guided by the local intrinsic dimensionality of the HD data to define seed locations for series of random walks; These random walks on the HD KNN graph are then visualized on 2D scatter plots, enabling the user to interactively query for patterns related to cell migration (in the spatial domain) as well as cell maturation (in the HD gene expression domain).*

**CCS Concepts**

• ***Human-centered computing*** → *Visual analytics;*

## 1. Introduction

Spatial transcriptomics (ST) data provides new opportunities to understand tissue biology by gene expression profiles and corresponding spatial information. This high-dimensional (HD), spatially resolved tissue imaging is primarily used for life-science discovery research. HD gene expression profiles are often projected to low-dimensional (LD) space for visual inspection of cell type compositions by dimensionality reduction (DR) techniques like PCA, t-SNE and UMAP [HWRB22]. DR embeddings give insight into data structure, showing clusters, transitions between clusters and generic trends in the data. Coloring the cell nodes in DR embedding with gene expression is commonly used to visualize the underlying pattern, but this is not able to reveal how cells are connected in HD space. ST data enables the exploration of cell-cell interactions and spatial cell migration trajectories. To characterize cell maturation, random walks and Gaussian Diffusion processes are widely used [LL21]. However the relationship between DR embedding, spatial location and HD gene expression profiles has not been fully explored. In addition, most computational strategies are script-based and lack interactive data exploration facilities with a direct feedback loop to the user.

In this work, we present SpaceWalker: an interactive visual analytics tool for HD gene expression profiles and spatial representations, enabling hypothesis generation in spatial biology of tissue development.

## 2. SpaceWalker

An overview of the proposed workflow is shown in Figure 1. The input of SpaceWalker is a cell-by-gene expression matrix, with spatial coordinates attached to the cell dimension [AMMR20]. SpaceWalker enables interaction in 2D representations of the data, combining similarity-based DR embeddings such as UMAP and t-SNE with spatial representations of the ST data.

The key idea behind SpaceWalker is to guide the exploration by features that express the local information complexity of the HD data space, as a proxy for biological variability. To this end, the local intrinsic dimensionality is computed for each data point in HD gene expression space and used to color the spatial map and 2D embedding, to steer the exploration by biological variability. When the user interacts with one point of interest on one of the 2D maps, genes with significant local spatial variability are selected by two spatial filters: 1) a peak filter, contrasting the gene expression vectors between two 2D neighborhoods with different sizes, and 2) a 2D gradient filter, which identifies genes with strong local edges.

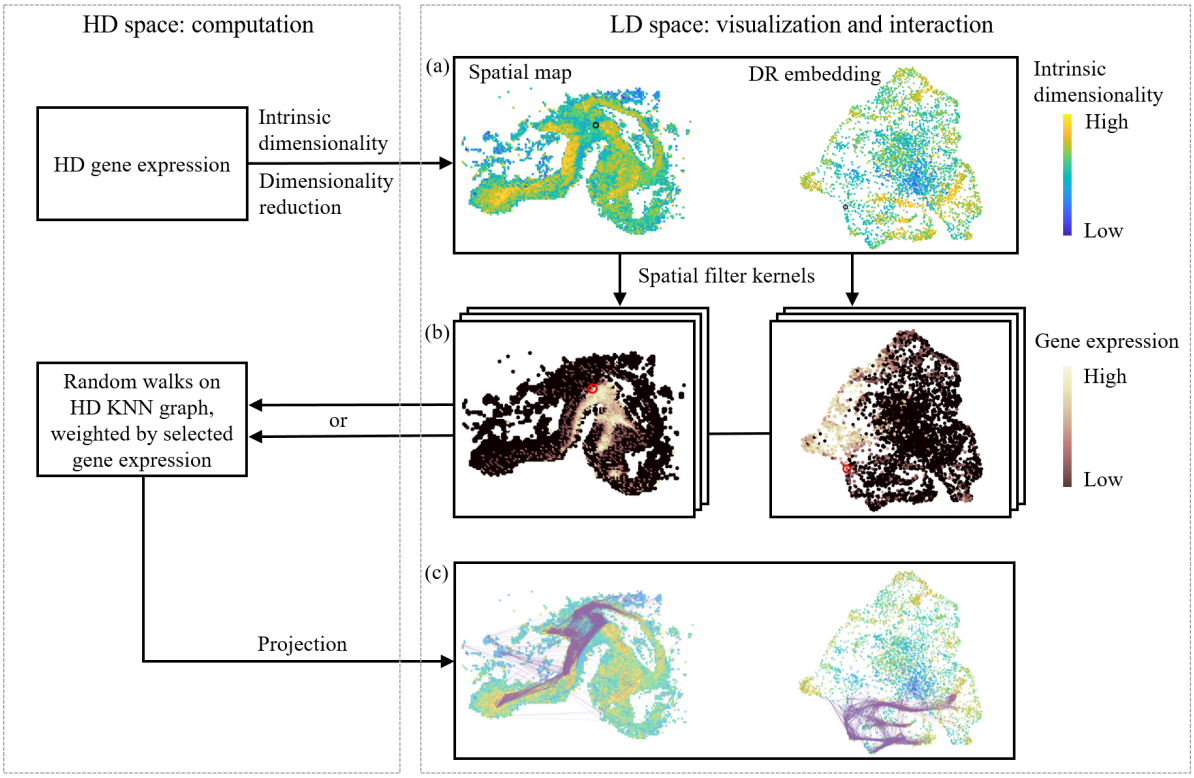


**Figure 1:** *Overview pipeline of the SpaceWalker tool. The user interacts with 2D scatter plots (a), where the spatial map and DR embedding are color-coded with the local intrinsic dimensionality of the HD data to highlight areas of potential interest. Genes with spatially varying local features (intensity peaks or gradients) are computed via spatial filter kernels, and presented to the user (b). Starting from a user selected node, a series of random walks is initiated to trace the structure of the KNN graph of the HD gene expression profiles, with expression values of a selected gene as weights in the random walks. Random walk paths are then projected back onto the scatterplots (c), revealing processes related to cell maturation and migration during neurodevelopment.*

Selected genes are presented in multiple complementary views to the user for visual inspection. For the gradient filter, the edge orientation of the selected gene can be shown as streamlines.

From the user-defined seed location, sets of random walks are conducted, to estimate the local topology of the HD space. Random walks are based on a KNN graph of the HD gene expression profiles. Each node in the graph represents a cell and graph edge weights are expression values of one selected gradient or peak gene identified by the spatial filters. By projecting the random walks in HD space to LD maps, we bring HD data structure and selected gene expression to LD representation together, illustrating the possible cell transition patterns. In Figure 1 (c), the random walk projections reveal known neurodevelopmental trajectories: radial glia migrating towards the midbrain to hindbrain region during cell maturation.

## 3. Future Work

We introduced SpaceWalker, a work-in-progress visual analytics approach for interactive ST data exploration. We aim to extend the idea of applying random walks in HD space and projecting back in LD space and plan to work with domain experts on a structure evaluation of the tool.

## 4. Acknowledgement

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860173.

## References

[AMMR20] ABDELAAL T., MOURRAGUI S., MAHFOUZ A., REINDERS M. J.: Spage: spatial gene enhancement using scrna-seq. *Nucleic acids research 48*, 18 (2020), e107–e107. 1

[HWRB22] HUANG H., WANG Y., RUDIN C., BROWNE E. P.: Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. *Communications biology 5*, 1 (2022), 1–11. 1

[LL21] LIU B., LI Y.: Analysis and visualization of spatial transcriptomic data. *Frontiers in Genetics* (2021), 2852. 1